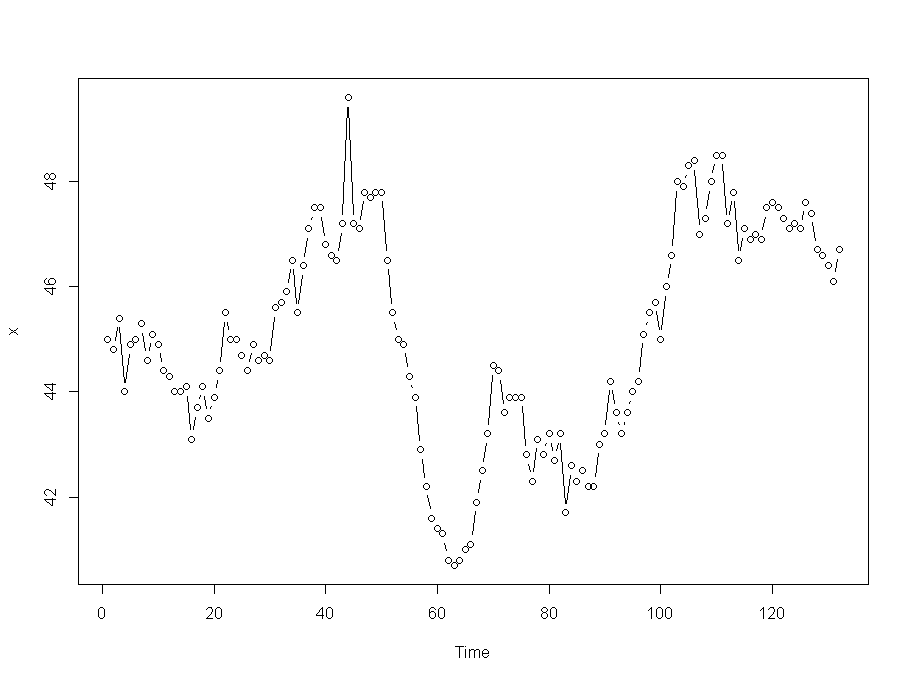
**Stat 510 Week 11 Homework**

**Submit answers in a Word document or a pdf to the designated ANGEL Drop Box in the Drop Boxes folder.**

1. Use the “metalsemploy.dat” dataset in the Week 11 folder. The data are proportional to monthly numbers of people employed in metals processing work in a state of the United States. There are *n* = 132 data values. The file can be read in R as x = scan ("metalsemploy.dat")

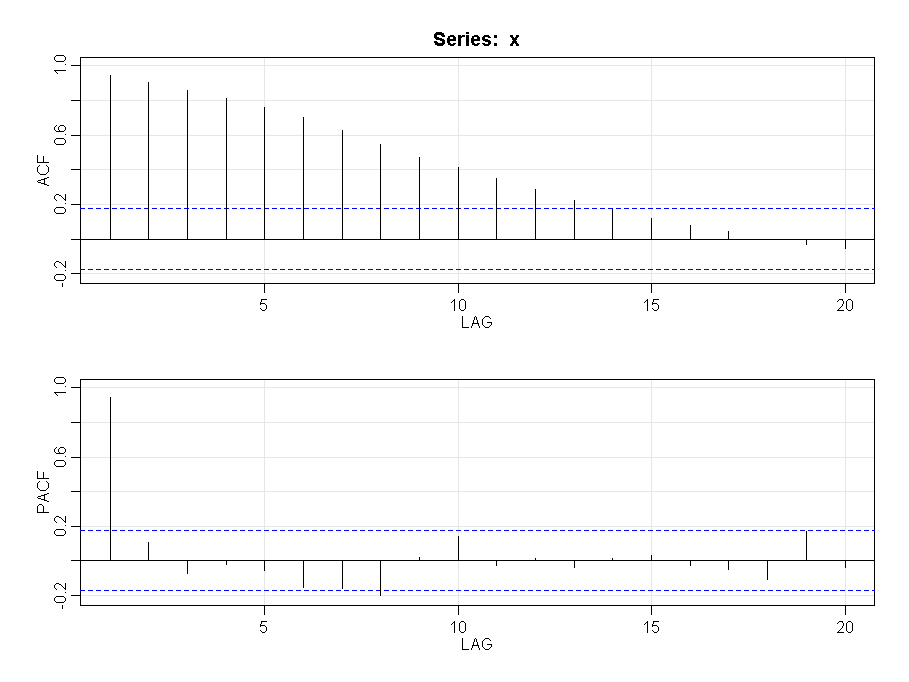
**a.** Create a time series plot of the series. Give the plot and a brief interpretation.



We note the following:

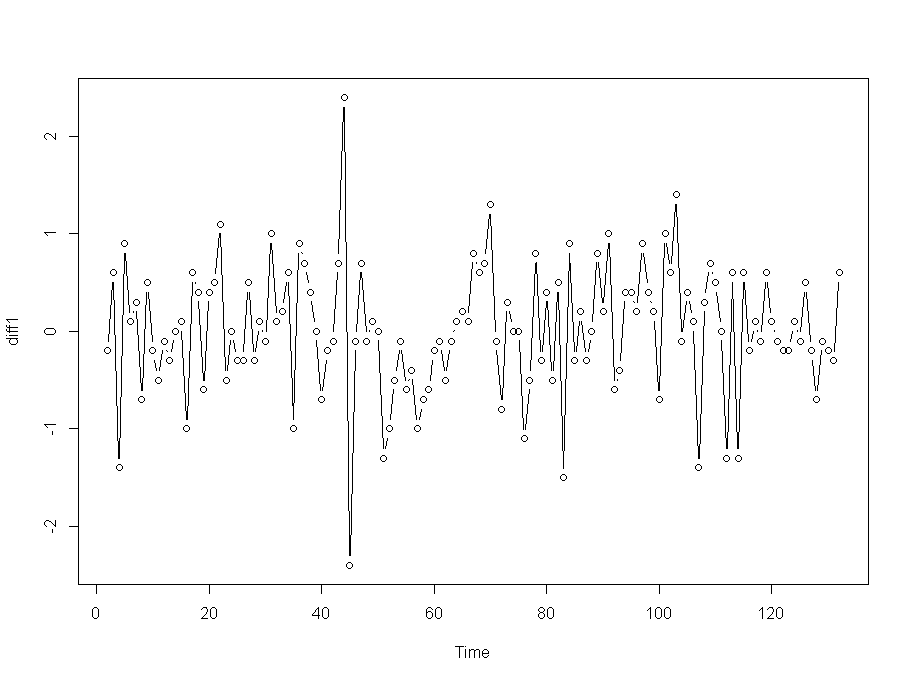
* There appears to be some trend in the series. The series gradually increases except for the big drop between 40 and 60
* It appears that the points at t very closely follow the values at t-1. This indicates a high value of correlation between xt and xt-1. We have to perform further analysis to confirm.
* The mean doesn’t appear to be constant.
* It appears that there are a few periods of increased variation, most notably between t=40 and t=70. However further analysis is required for that.

**b**. Determine the ACF and PACF of the series. Write a brief interpretation.



We see here that ACF trails to 0 and PACF spikes at lag=1. This is typical structure for AR(1) model. We also see that the correlation is very high (almost 1) and therefore we may want to look at differences.

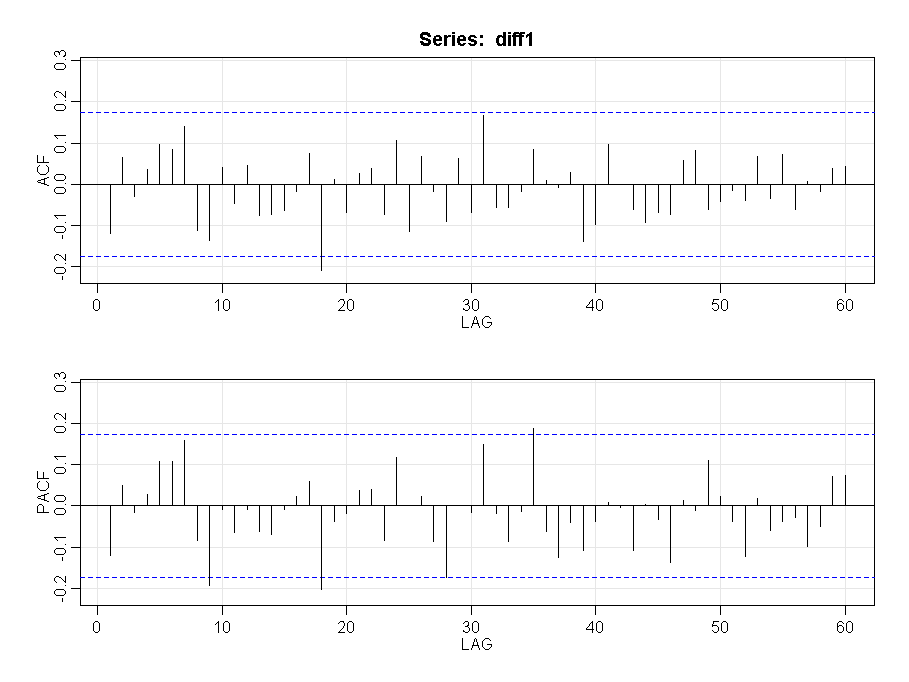
**c.** Here’s a hint to parts A and B. The series slowly oscillates, almost in a wandering fashion, and the first lag autocorrelation is very close to 1. This suggests a possible random walk or . Take first differences of the series and create a time series plot of the series of first differences. Give the plot and write a brief interpretation. (Hint: See Lesson 11.1 for this week.)



The time series has the following properties:

* There are a few periods of changed variation, most notably bet t= 40 and t = 60.
* There is no consistent trend (upward or downward) over the entire time span. The mean seems to be stationary now.

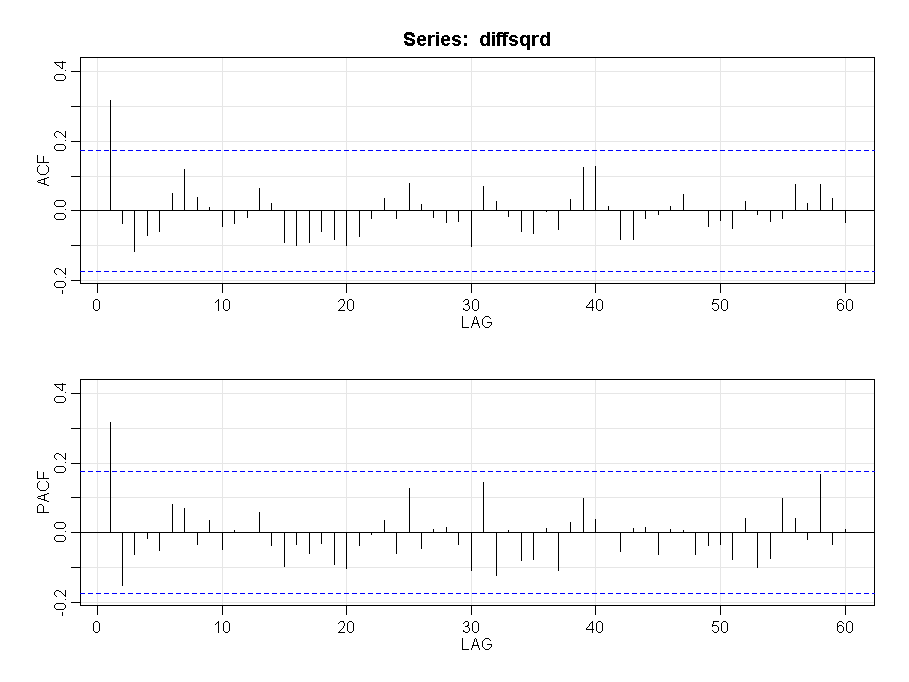
**d**. Determine the ACF and PACF of the first differences. What model (if any) is suggested?



Note that no correlations are significant, so the series looks to be white noise. This is in alignment with the time series plots.

**e**. Square the first differences. In R this can be done with a command like diffsqrd = diff1\*diff1.

(Exact command will depend upon what you called the first differences.) Determine the ACF and PACF of the squared first differences. What model (or models) is suggested?



Here we see that there is a peak in both ACF and PACF at lag 1. We should look at using the models ARCH(1) and GARCH(1,1)

**f.** Parts d and e suggest the possibility of an ARCH(1) model for the first differences. One way to estimate the parameters is to utilize the fact that the coefficients for an AR(1) model for the squared values (differences here) are the same as for the ARCH(1) model for the variance. Thus we can fit an AR(1) model to the squared differences to estimate the parameters of the ARCH(1) variance equation. Fit an AR(1) model to the squared differences. What are the estimated coefficients? (Intercept and “slope”.) Note: If you use arima, it mistakenly labels the mean as the intercept. If you use sarima, it correctly labels the mean. You must then calculate intercept = mean \*(1-).

An AR(1) on the squared differences gives:

ttable

Estimate SE t.value p.value

ar1 0.3150 0.0825 3.8164 2e-04

xmean 0.4384 0.0990 4.4275 0e+00

The estimated model can be written as (xt - 0.4384) = 0.3150 (xt-1 – 0.4384) + wt

This is equivalent to xt = 0.4384 – (0.3150 \* 0.4384) + 0.3150 xt-1 + wt

xt = 0.3003 + 0.3150 xt-1 + wt

Intercept = 0.3003 and Slope = 0.3150

Intercept can also be calculated as 0.4384 \* (1 – 0.3150) = 0.3003

We can also see the output from arima:

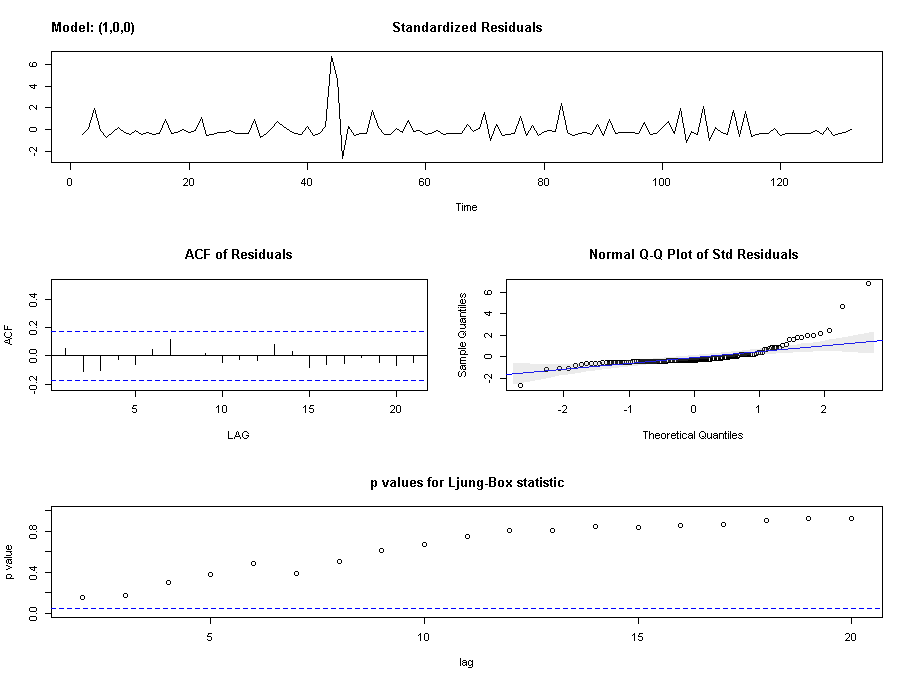
Coefficients:

ar1 intercept

0.3150 0.4384

s.e. 0.0825 0.0990

Here we see that the mean is mislabeled as intercept.



**g**. Part f generated estimates of α0 and α1 for the variance model . Estimate the variance at time *t* = 133. At time 132 (end of the series), the value of = 46.7 and at time 131, =46.1. Thus, the first difference at time 132 is = 0.6. (Note that the length of the first differences in R is 131, however the start parameter is set to 2 and so the 131st entry of the differences vector is actually the first difference at t=132.)

Var (yt | yt−1) = σ2t = 0.3003 + 0.3150 y2t−1

= 0.6 therefore 2 = 0.36

Therefore Var(y133 | y132) = σ2t = 0.3003 + 0.3150 y2t−1 = 0.3003 + (0.3150 \* 0.36) = 0.4137

2. Use the files y1hw11.2.dat, y2hw11.2.dat, and y3hw11.2.dat from the Week 11 folder. The files simulate food and beverage price index values for (the same) 400 consecutive months at commodities exchanges in 3 different US cities.

Refer to Lesson 11.2 of this week for guidance on R code for this problem.

A. Estimate the coefficients of a VAR(1) model for the three series. Check the significance of the intercept and trend; include just what is necessary in your final analysis by searching the help file for the VAR command (?VAR). Write the three resulting regression equations (one for each series).

Estimation results for equation y1:

===================================

y1 = y1.l1 + y2.l1 + y3.l1 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 2.667e-01 4.704e-02 5.669 2.77e-08 \*\*\*

y2.l1 -1.498e-01 4.370e-02 -3.428 0.000671 \*\*\*

y3.l1 1.771e-01 4.072e-02 4.348 1.75e-05 \*\*\*

const 1.881e+02 2.322e+01 8.101 6.90e-15 \*\*\*

trend 1.302e-04 4.304e-04 0.303 0.762361

We find for y1:

* Trend is not significant
* The equation is: yt,1Hat = 188.1 + 0.2667 yt-1,1 – 0.1498 yt-1,2 + 0.1771 yt-1,3

Estimation results for equation y2:

===================================

y2 = y1.l1 + y2.l1 + y3.l1 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 -1.654e-01 5.281e-02 -3.131 0.00187 \*\*

y2.l1 7.810e-02 4.906e-02 1.592 0.11216

y3.l1 1.953e-01 4.572e-02 4.273 2.43e-05 \*\*\*

const 1.447e+02 2.607e+01 5.549 5.27e-08 \*\*\*

trend -7.579e-06 4.832e-04 -0.016 0.98749

We find for y2:

* Trend is not significant
* The equation is: yt,2Hat = 144.7 – 0.1654 yt-1,1 + 0.0781 yt-1,2 + 0.1953 yt-1,3

Estimation results for equation y3:

===================================

y3 = y1.l1 + y2.l1 + y3.l1 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 2.245e-01 4.970e-02 4.516 8.34e-06 \*\*\*

y2.l1 3.785e-01 4.617e-02 8.198 3.48e-15 \*\*\*

y3.l1 2.871e-01 4.303e-02 6.674 8.49e-11 \*\*\*

const 1.625e+02 2.453e+01 6.622 1.16e-10 \*\*\*

trend 4.096e-04 4.548e-04 0.901 0.368

We find for y3:

* Trend is not significant
* The equation is: yt,3Hat = 162.5 + 0.2245 yt-1,1 + 0.3785 yt-1,2 + 0.2871 yt-1,3

B. Estimate the coefficients of a VAR(2) model for the three series; include both, constant, trend, or neither as necessary. Write the equation for City 3 only.

Estimation results for equation y1:

===================================

y1 = y1.l1 + y2.l1 + y3.l1 + y1.l2 + y2.l2 + y3.l2 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 2.742e-01 5.070e-02 5.408 1.11e-07 \*\*\*

y2.l1 -1.571e-01 4.530e-02 -3.468 0.000583 \*\*\*

y3.l1 2.002e-01 4.783e-02 4.185 3.52e-05 \*\*\*

y1.l2 -5.665e-02 5.101e-02 -1.111 0.267457

y2.l2 -3.810e-02 4.810e-02 -0.792 0.428789

y3.l2 6.248e-03 4.479e-02 0.139 0.889135

const 1.999e+02 2.694e+01 7.420 7.38e-13 \*\*\*

trend 1.196e-04 4.337e-04 0.276 0.782839

Estimation results for equation y2:

===================================

y2 = y1.l1 + y2.l1 + y3.l1 + y1.l2 + y2.l2 + y3.l2 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 -1.640e-01 5.694e-02 -2.881 0.00419 \*\*

y2.l1 6.517e-02 5.087e-02 1.281 0.20092

y3.l1 2.048e-01 5.372e-02 3.813 0.00016 \*\*\*

y1.l2 -5.070e-02 5.729e-02 -0.885 0.37674

y2.l2 -2.597e-02 5.402e-02 -0.481 0.63101

y3.l2 3.511e-02 5.031e-02 0.698 0.48563

const 1.487e+02 3.026e+01 4.916 1.31e-06 \*\*\*

trend -1.511e-05 4.871e-04 -0.031 0.97527

Estimation results for equation y3:

===================================

y3 = y1.l1 + y2.l1 + y3.l1 + y1.l2 + y2.l2 + y3.l2 + const + trend

Estimate Std. Error t value Pr(>|t|)

y1.l1 2.252e-01 5.362e-02 4.199 3.32e-05 \*\*\*

y2.l1 3.903e-01 4.791e-02 8.147 5.10e-15 \*\*\*

y3.l1 2.987e-01 5.059e-02 5.905 7.69e-09 \*\*\*

y1.l2 2.614e-02 5.395e-02 0.484 0.628

y2.l2 -8.143e-03 5.087e-02 -0.160 0.873

y3.l2 -4.393e-02 4.737e-02 -0.927 0.354

const 1.671e+02 2.849e+01 5.863 9.71e-09 \*\*\*

trend 4.418e-04 4.587e-04 0.963 0.336

We find for city 3:

* Trend is not significant. Only 1st lags are significant.
* The equation is: yt,3Hat = 167.1 + 0.2252 yt-1,1 + 0.3903 yt-1,2 + 0.2987 yt-1,3
* The above equation is almost equal to the equation for VAR(1) model

C. Use formula 5.90 at the top of page 306 in the text to calculate the BIC value for the VAR(1) model and for the VAR(2) model. Give both values. On the basis of these two BIC values, which model is better? (Lower is better.)

Here’s the format of R commands for the VAR(1) model:

n = 400

k = 3

p = 1

log(det(summary(fitvar1)$covres)) + (k^2\*p\*log(n)/n)

In the formula, k = number of variables and n = sample size, so those will stay the same from model to model. The “fitvar1” within the summary command is the object name that you specify in the VAR(1) fit. That will have to change for the VAR(2) fit (as does the value of p, the order of the AR).

For the VAR(1) model:

BIC\_var1

[1] 0.3919035

For the VAR(2) model:

BIC\_var2

[1] 0.5452319

Therefore VAR(1) model is better.

D. For the VAR(1) model, examine the ACF of the residuals for each series. Discuss whether each is white noise (completely random) or not.

We have:

* The ACF of the residuals for the 3 series are plotted below. They resemble white noise.

